



## Edinburgh Research Explorer

### **Mosaic structure of the human immunodeficiency virus type 1 genome infecting lymphoid cells and the brain**

**Citation for published version:**

Morris, A, Marsden, M, Halcrow, K, Hughes, ES, Brettle, RP, Bell, JE & Simmonds, P 1999, 'Mosaic structure of the human immunodeficiency virus type 1 genome infecting lymphoid cells and the brain: evidence for frequent in vivo recombination events in the evolution of regional populations', *Journal of Virology*, vol. 73, no. 10, pp. 8720-31.

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Journal of Virology

**Publisher Rights Statement:**

Copyright © 1999, American Society for Microbiology

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Mosaic Structure of the Human Immunodeficiency Virus Type 1 Genome Infecting Lymphoid Cells and the Brain: Evidence for Frequent In Vivo Recombination Events in the Evolution of Regional Populations

A. MORRIS,<sup>1</sup> M. MARSDEN,<sup>1</sup> K. HALCROW,<sup>1</sup> E. S. HUGHES,<sup>1</sup> R. P. BRETTE,<sup>2</sup>  
J. E. BELL,<sup>3</sup> AND P. SIMMONDS<sup>1\*</sup>

*Department of Medical Microbiology, University of Edinburgh, Edinburgh EH8 9AG,<sup>1</sup> and Regional Infectious Diseases Unit, Western General Hospital,<sup>2</sup> and Department of Neuropathology, University of Edinburgh, Western General Hospital,<sup>3</sup> Edinburgh EH4 2XU, United Kingdom*

Received 16 December 1998/Accepted 7 July 1999

**In addition to immunodeficiency, human immunodeficiency virus type 1 (HIV-1) can cause cognitive impairment and dementia through direct infection of the brain. To investigate the adaptive process and timing of HIV-1 entry into the central nervous system, we carried out an extensive genetic characterization of variants amplified from different regions of the brain and determined their relatedness to those in lymphoid tissue. HIV-1 genomes infecting different regions of the brain of one study subject with HIV encephalitis (HIVE) had a mosaic structure, being assembled from different combinations of evolutionarily distinct lineages in *p17<sup>gag</sup>*, *pol*, individual hypervariable regions of *gp120* (V1/V2, V3, V4, and V5), and *gp41/nef*. Similar discordant phylogenetic relationships were observed between *p17<sup>gag</sup>* and V3 sequences of brain and lymphoid tissue from three other individuals with HIVE. The observation that different parts of the genome of HIV infecting a particular tissue can have different evolutionary histories necessarily limits the conclusions that can be drawn from previous studies of the compartmentalization of distinct HIV populations in different tissues, as these have been generally restricted to sequence comparisons of single subgenomic regions. The complexity of viral populations in the brain produced by recombination could provide a powerful adaptive mechanism for the spread of virus with new phenotypes, such as antiviral resistance or escape from cytotoxic T-cell recognition into existing tissue-adapted virus populations.**

Isolates of human immunodeficiency virus (HIV) show great heterogeneity in their kinetics of replication, coreceptor usage, cellular tropism, and cytopathic effects. Differences in these properties have been hypothesized to underlie in vivo differences in pathogenicity, which in turn may influence the rate of disease progression in HIV-infected individuals (1, 2, 5, 9, 11, 16, 19, 20, 28, 40, 44). Differences in phenotype may also underlie differences in in vivo cellular tropism, which would substantiate the hypothesis that the different populations of HIV infecting lymphoid tissue, brain, and other tissues may have originated through an adaptive process following primary infection. Difficulties in recovering HIV from nonlymphoid tissues have to date prevented extensive analysis of their biological properties, although the existence of consistent sequence differences in the envelope gene from those recovered from lymphoid tissues provides indirect evidence for specific cellular tropisms (3, 13, 14, 17, 22, 26, 27, 30, 31, 34–36, 42). However, an alternative hypothesis proposes that differences in the rate of virus turnover in different cell types may lead to the observed population differences, given the rapid temporal change in HIV populations over time in peripheral blood mononuclear cells, in lymph nodes (LN), and among HIV variants recovered from the gastrointestinal tract (27, 45, 46).

To investigate whether differences in V3 and elsewhere in the HIV genome reflect tissue adaptation, or whether they

arise simply through limited spatial or temporal sampling, we have compared nucleotide sequences in different regions of the HIV type 1 (HIV-1) genome from lymphoid tissue with autopsy samples from anatomically separated parts of the brains of four study subjects with HIV encephalitis (HIVE). We also determined whether sequences in *p17<sup>gag</sup>*, *pol*, and different regions of the *env* gene (V1/V2, V3, V4, V5, and *gp41/nef*) between different brain samples and those from lymphoid tissue provided equivalent evidence for tissue-specific compartmentalization of HIV-1.

## MATERIALS AND METHODS

**Study subjects.** Frozen samples of LN or spleen and from several anatomically distinct regions of the brain from four individuals with HIV giant cell encephalitis were stored at autopsy (risk group, CD4 counts, and brain pathology are summarized in Table 1). Subject NA129 had received zidovudine monotherapy for 17 months up to approximately 1 year before death; ddC was used for 1 month, finishing 3 months before death. The other study subjects had received minimal antiviral treatment: for NA234, a single course of zidovudine for a duration of 1 month at 1 year before death; for NA021, zidovudine intermittently over 1 year at 5 years before death and zidovudine-ddC for 1 month at 1 year before death; for NA173, zidovudine for 4 months at 2 years before death and then for 1 month at 1 year before death. None of the study subjects showed evidence for genotypic resistance to zidovudine or other antiviral agents (42a).

**Pathology examination.** The brains were examined pathologically as previously described (4). Assessment of pathology findings was undertaken blind to the PCR analysis and validated by three independent observers.

**DNA extraction and amplification.** DNA was extracted from the brain, LN, and spleen and quantified as previously described (15). Total DNA was quantified spectroscopically, while HIV proviral DNA was semiquantitated by amplification, using the *p17<sup>gag</sup>* primers, of serial 10-fold dilutions of DNA, with the last positive dilution used to indicate the minimum proviral load in the sample. Samples were used for sequence comparison only if proviral frequencies were >100 copies/10<sup>6</sup> cells, and this excluded analysis of left parietal (LP) and both cerebellum samples from NA234. Low levels were detected in two atrophied LN

\* Corresponding author. Mailing address: Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom. Phone: 44 131 650 3138. Fax: 44 131 650 6531. E-mail: Peter.Simmonds@ed.ac.uk.

TABLE 1. Clinical background and pathology appearance of study subjects

Study subject	Age (yr)	Risk group <sup>a</sup>	CD4/ $\mu$ l		HIV-related pathology <sup>b</sup>						Other pathology
			1 yr before death	3 mo before death	PV	WM	GM	BG	BS	Cere	
NA173	34	MH	1	0	++	+/-	+/-	+/-	+/-	-	Toxoplasma
NA021	49	IVDU	300	87	+	++	+	+	+/-	+/-	
NA128	48	MH	12	4	+/-	+++	++	++	++	-	Toxoplasma
NA234	28	IVDU	90	30	+	++	++	++	+	+	Lymphoma

<sup>a</sup> MH, male homosexual; IVDU, intravenous drug user.

<sup>b</sup> Scored as - (negative), +/- (sparse pathology), +, (slight pathology), or +++ (severe pathology in different brain compartments). PV, perivascular; WM, white matter; GM, grey matter; BG, basal ganglia; BS, brain stem; Cere, cerebellum.

samples from NA128, and lymphoid sequences were therefore obtained from the spleen. For nucleotide sequencing, 1- $\mu$ g aliquots of extracted DNA were amplified as previously described (39) or amplified after dilution to an endpoint (see below).

Amplified DNA was either sequenced directly by cycle sequencing (Amersham) or cloned into pGEM, using poly(T) overhangs (pGEM-T Vector system; Promega). Miniprep DNA from clones was sequenced by using a Sequenase version 2.0 kit (U.S. Biochemical) according to the manufacturer's protocol.

**Nucleotide sequencing.** For samples with evidence of sequence heterogeneity, amplified DNA was cloned and approximately 10 clones were sequenced. Sequences from other samples without evidence for heterogeneity (no detectable multiple bands at any position on a sequencing gel lane) or length polymorphism analysis (23) were directly sequenced. The sequence data set for the p17<sup>gag</sup> region of NA234 was assembled from cloned sequences from each of the brain regions and extended from positions 405 to 795 in the HIV<sub>LAI</sub> genomic sequence (GenBank accession no. K02013). Sequences in the *pol* region of NA234 were directly sequenced from amplified DNA and compared between bases 2189 and 2842, using the PCR primers used for detection of antiviral resistance mutations (43); there was no evidence for intrasample sequence heterogeneity. Sequences in the V1/V2 region from NA234 were amplified by using primers as previously described (23). Amplified DNA was homogeneous by length polymorphism analysis, and nucleotide sequences from regions other than LN, left frontal (LF), and right frontal (RF) were obtained by direct sequencing of amplified DNA. Sequences in V3 and V4/V5 were amplified by using primers as previously described (38). Length polymorphisms in the V4/V5 region in some brain samples from NA234 necessitated multiple clones to be sequenced. Sequences from each tissue were amplified in the gp41/*nef* region by using outer primers 5'-AGGGC TGCTATTAACAAGAGATGG-3' and 5'-GTAGCTGCTGTATTGCTACTT GTG-3' and inner primers 5'-AGGTCTTCAGACCTGGAGGAGG-3' and 5'-TATTGCTACTTGTGATTGCTCCATG-3' (5' base positions 7166, 8541, 7215, and 8531, respectively). Direct sequencing was carried out on the amplified DNA from the 3' end by using the antisense inner primer, while an internal primer (5'-TGAGGGGACAGATAGGGTTATAGA-3', position of 5' base 8287 in HIV<sub>LAI</sub> genome) was used to sequence the 5' end.

Sequences were aligned and distances were estimated with the Simmonic 2000 Sequence Editor package. Synonymous and nonsynonymous distances and standard errors were estimated by the method of Nei and Gojobori (33). Phylogenetic analysis was carried out with the MEGA program (29). The nucleotide sequences from p17<sup>gag</sup> and V3 amplified from each of the study subjects were compared with each other and with a range of standard HIV-1 variants. Each set of sequences from the four study subjects was monophyletic in both genomic regions and distinct from those of the published sequences of subtype B: HIVSF2 (K02007), HIVRF (M17451), HIVOYI (M26727), HIVLAI (K02013), HIVJRFL (M74978), HIVYU2 (M93258), HIVCAM1 (D10112), HIVNYSOG (M38431), HIVHAN (U43131), HIVWMJ22 (M12507), and HIVSFAAA (M65024). This comparison provided no evidence for coinfection with more than one epidemiologically unrelated HIV strains or for intersample or exogenous laboratory contamination.

**Nucleotide sequence accession numbers.** Nucleotide sequences obtained in this study have been submitted to GenBank and assigned accession no. AF174692 through AF175123.

## RESULTS

**Sequence relationships in different regions of the HIV-1 genome.** Autopsy samples were obtained from multiple regions of the brain and from LN of an individual (NA234) with autopsy evidence of HIV. The p17<sup>gag</sup> region was amplified by nested PCR and cloned from samples with virus loads of greater than 100 proviral copies/10<sup>6</sup> cells. These nucleotide sequences formed two distinct evolutionary lineages, each with bootstrap support of  $\geq 80\%$ . Sequences from LN, brain stem,

right parietal (RP) and LF regions (lineage A) showed a closer sequence relationship to each other than to variants obtained from elsewhere in the brain (lineage B) (Fig. 1). At synonymous sites, the mean pairwise Jukes-Cantor (J-C) distances were 0.057 (mean standard error  $\pm 0.026$ ) among members of lineage A and 0.055 ( $\pm 0.024$ ) for lineage B, almost nonoverlapping with the range of pairwise distances observed between the two lineages (mean synonymous distance  $0.109 \pm 0.038$ ). On the basis of the previously established mean rate of sequence change in this region of *gag* (0.6 to 0.7% per site per year [24, 25]), these distances suggest a time of divergence between lineages A and B of around  $8.4 \pm 2.9$  years, compared with a likely duration of HIV infection of around 10 years in this study subject.

In contrast to LN sequences, those from other regions of the brain were relatively homogeneous, apart from those from the RP region, which formed at least three separate clusters within lineage A, and a single sequence from the LF region, which grouped in lineage B instead of lineage A. Sequences in lineage B, comprising exclusively brain-derived variants, clustered by tissue origin, with a tendency for variants recovered from adjacent tissues to be more similar to each other (e.g., left occipital [LO] and right occipital [RO] regions) than to variants from virus with greater physical separation (e.g., occipital region to RF region). Most regional variants in lineage B had common ancestors distinct from any variants found in lymphoid tissue, and more recent times of divergence than that between lineages A and B can be inferred. Variants within lineage B may therefore have originated from the spread of HIV within the brain rather than from multiple seeding from the peripheral circulation and thus differ in origin from brain-derived variants in lineage A (LF and RP). To confirm the separate groupings of LF and RF sequences, DNAs extracted from these tissues and from LN were separately analyzed by limiting dilution as previously described (39). Sequences obtained were similar or identical to those obtained by cloning of PCR products (data not shown).

In marked contrast to the relationships observed in p17<sup>gag</sup>, LF and RF sequences from both V1/V2 and V3 regions in gp120 grouped together in a distinct lineage from those of LN sequences (bootstrap support  $\geq 80\%$ ) (Fig. 2A and B), demonstrating that variants from LF and RF regions shared a common ancestor distinct from that of variants in the LN region. The probability of this discordant phylogeny arising through sampling was  $<0.0001$  (Fisher's exact test). To investigate the robustness of the difference in branching order in p17<sup>gag</sup> and V3, a user-defined tree of p17<sup>gag</sup> sequences with the branching order of V3 was created by using RETREE in the PHYLIP package (18). With the Hasegawa-Kishino-Yano test, a p17<sup>gag</sup> tree with the V3 branching order was significantly less

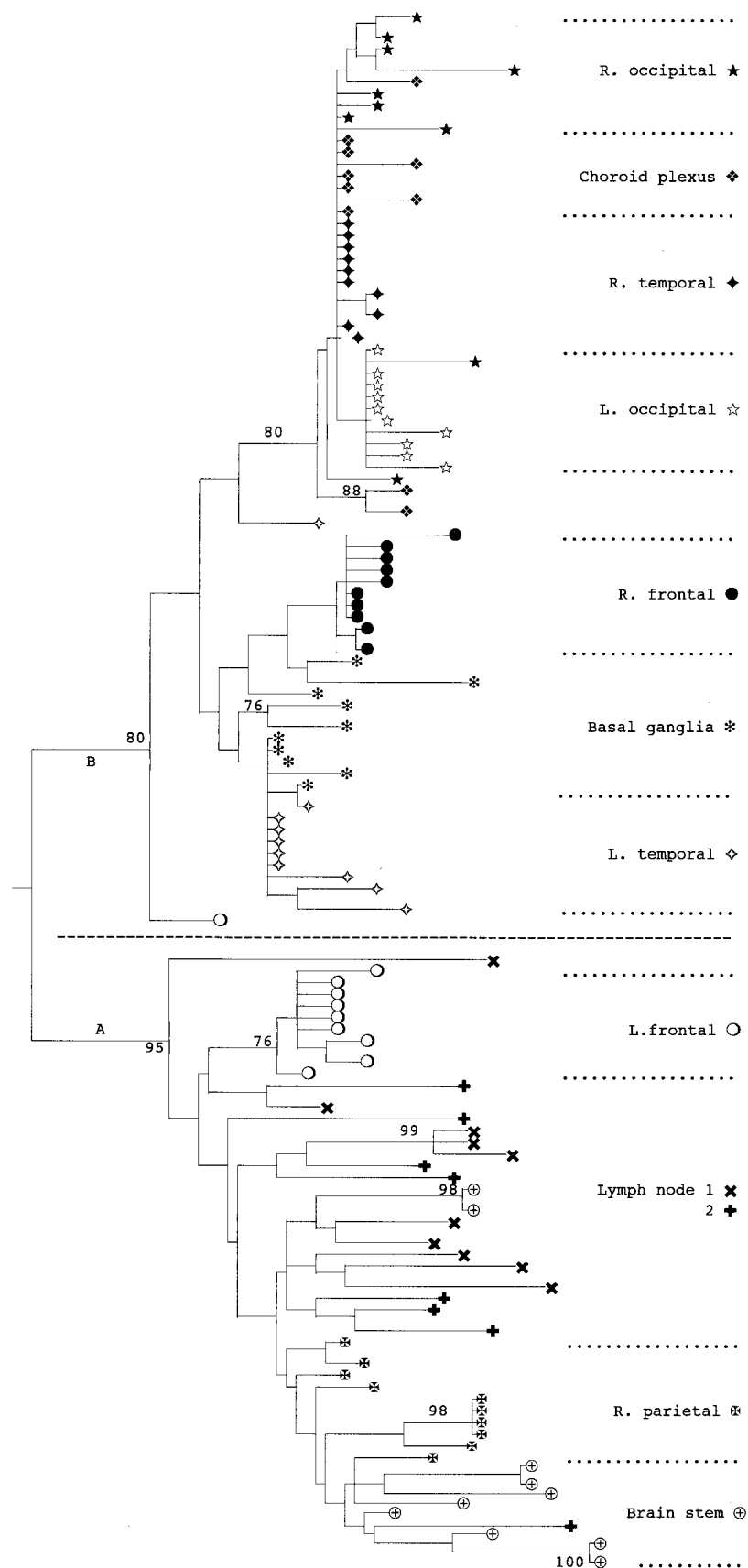


FIG. 1. Phylogenetic analysis of the p17<sup>gag</sup> region amplified from different regions of the brain and from lymphoid tissue of NA234. Divergence between nucleotide sequences was estimated from J-C distances (scale indicated below tree), and the tree was constructed from the distance matrix by the neighbor-joining method. The robustness of groupings was indicated by bootstrap resampling of 100 data sets, with values of  $\geq 75\%$  indicated on branches; lineages are indicated by single letters. The tree was rooted by using the sequence of HIV<sub>LAI</sub> (accession no. K02013) as an outgroup.

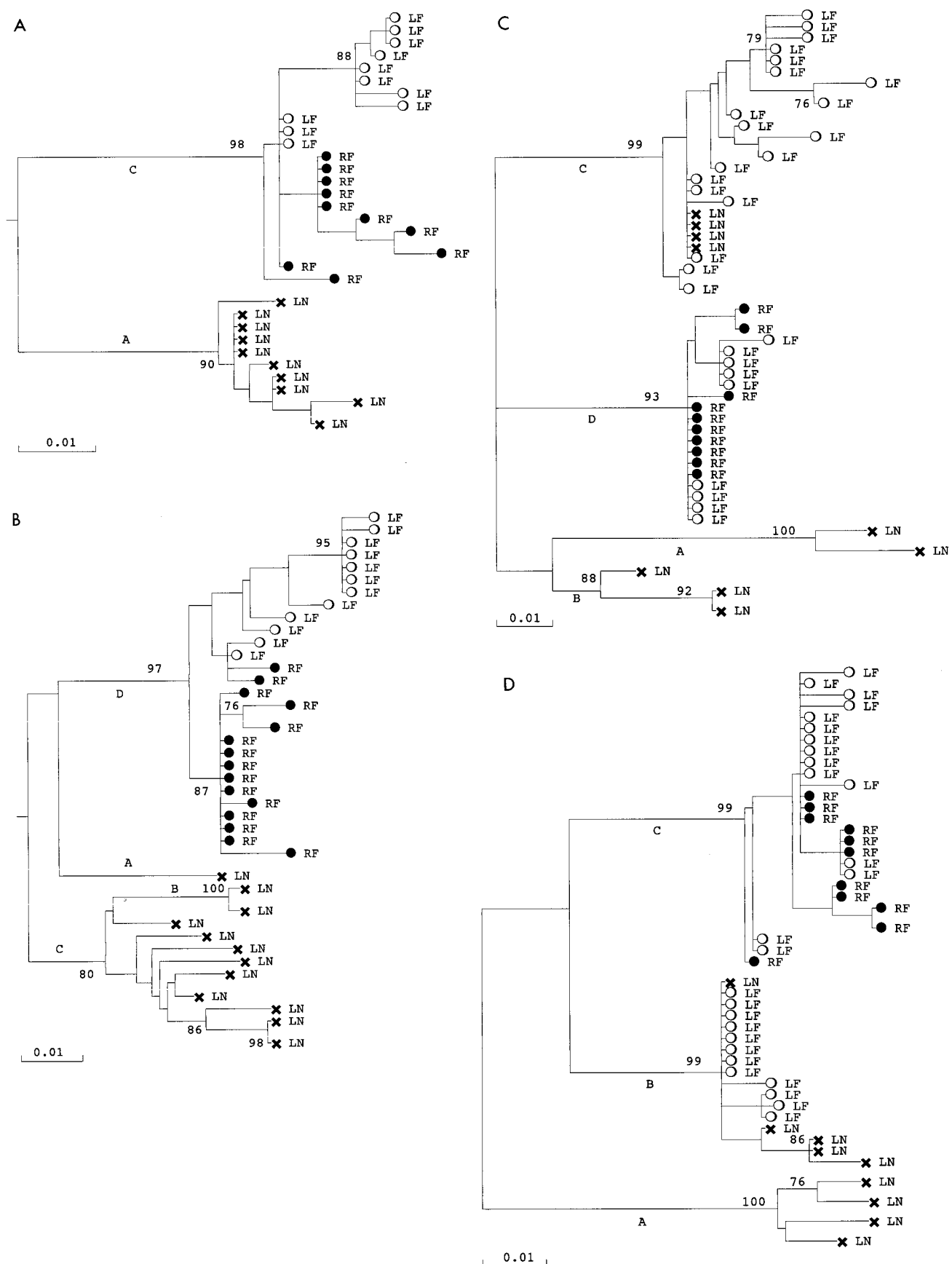


FIG. 2. Comparison of sequences from LF and RF regions of the brain of NA234 with those of the LN region in V1/V2 (A), V3 (B), V4 (C), and V5 (D) hypervariable regions (outgroup, symbols, and bootstrap method as for Fig. 1).

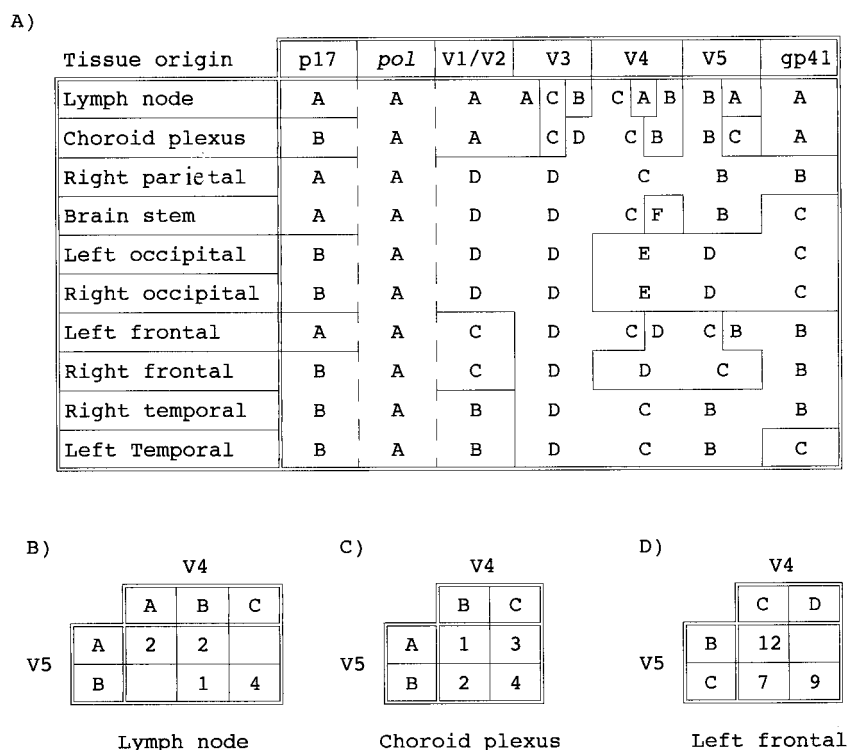


FIG. 3. (A) Distribution of phylogenetic groupings in different regions of the genome of HIV-1 variants infecting NA234. Letters refer to bootstrap-supported groupings identified by phylogenetic analysis (Fig. 1, 2, and 3); vertical bars indicate discontinuities in groupings that require the minimum number of recombination events. The sequence conservation and monophylogeny of the *pol* region originated either through recombination or from a lower rate of sequence change compared with the rest of the genome; sequence relationships are indicated with an interrupted line to indicate this uncertainty. Three samples contained sequences distributed in more than one clade in V4 and V5, in the combinations tabulated in LN (B), CP (C), and LF (D) regions of the brain.

likely than the most likely tree calculated by using maximum likelihood (DNAML;  $P = 0.018$ ).

In the V4/V5 region, sequence relationships were more complex and also discordant from those observed both for p17<sup>gag</sup> and for V1/V2 or V3 (Fig. 2C, 2D, 3C, and 3D). Variants from the LN region formed three lineages (A, B, and C), with a fourth lineage comprising variants from the LF and RF regions (lineage D). However, approximately half of the LF variants grouped with lineage C. Similar sequence relationships were observed for the V5 region, with three main lineages comprising LN (A), LN and LF (B), and LF and RF (C). Because V4 and V5 sequences were amplified in the same PCR fragment, it was also possible to observe different combinations of lineages between the two regions (Fig. 3D). For example, V4 sequences of 12 LF clones were in lineage C in V4 and in lineage B in V5 (i.e., CB), while 9 clones were DC and 7 were CC. Similar reassortments were observed in LN sequences (one AA, two AB, one BB, and three BC).

To investigate further the tissue distribution of variants in different genomic regions, we obtained sequences from the *pol*, V1/V2, V3, and gp41/*nef* regions of variants recovered from other parts of the brain (Fig. 4). The most conserved region was the *pol* region, in which sequences from different regions of the brain did not display any significant phylogenetic groupings. The mean pairwise distance between *pol* sequences from different samples was 0.013 (0.048 at synonymous sites), lower than observed between variants within either lineage A or lineage B in the p17<sup>gag</sup> region. Either there was a lower rate of sequence change at synonymous sites in this part of the genome or the *pol* region of the genome originated after the

diversification of p17<sup>gag</sup>, implying the existence of recombination between the *pol* region and more variable regions of the genome.

In V1/V2, we observed four bootstrap-supported lineages, differing from each other by distances of 0.022 to 0.076, compared with distances of 0 to 0.008 within lineages. The distribution of sequences from different regions of brain did not match the distribution of sequences into the two lineages in p17<sup>gag</sup> (Fig. 3 and 4). For example, sequences from LN (A in p17<sup>gag</sup>) and choroid plexus (CP) (B in p17<sup>gag</sup>) regions were found in the same lineage in V1/V2, while sequences from LF and RP regions, which grouped with the LN sequence in p17<sup>gag</sup>, were on three separate lineages in V1/V2 (A, C, and D).

Similar complexity was observed among sequences from V3, V4, V5, and gp41/*nef*, with variants from the CP showing the greatest diversity in V3, V4, and V5. A subpopulation of CP variants grouped with those found in lymphoid tissue (e.g., lineages C in V3, B in V4 and V5, and A in gp41), but as described above for LN and LF, there were inconsistent relationships in the V4 and V5 regions (Fig. 3C), with all four combinations of lineages B and C in V4 and A and B in V5 being observed. Overall, no consistent relationship between lineages was observed among the other samples collected from the brain (Fig. 3), and apart from LO and RO regions, each region of the brain contained a different combination of phylogenetically distinct subgenomic fragments.

**Tissue-specific grouping.** Variants from different regions of the brain of NA234 consistently grouped separately from those recovered from LN in the V1/V2, V3, and gp41/*nef* sub-

## A) V1/V2

124

216

A	CP LN	C 10	PLCVTLKCKD VGNTTNITSS SWEKMEKGEI KNCSFKVTTT IKDKTQIEHA LFYKLDVVPF DNDNDNTSYR LISCNTSVIT QACPKVSFE
B	RT LT	C	.....R.T. .R.N..T. ....NI.. .S..V.K. ....N.....
C	LF RF	11 10	.....R.T. .R.N..T..N .....NI.. .S..V.K. ....N.....
D	BS LO RO RP	C C C C	.....T. .R.N..T. ....NI.. .S..V.Q. ....N.-- .....TN .R.N..T. ....NI.. .S..V.Q. ....N.-- .....T. .R.N..T. ....NI.. .S..V.Q. ....N.-- .....T. .R.N..T. ....NI.. .S..V.Q. ....N.--

## B) V3

276

357

A	LN	1	VIRSENFNN AKIIIVHLNK SIAINCTRPN SNTRKGIHIA PGRAFYATGD IIGDIRQAH NLSKAEWENT LKQVAMKRE Q
B	LN	2	.....T..... .V..H..... N..... .....I.R.D.N. ....I..K .
C	LN	9	.....T...Q.T. PVV..... N..... .R.D.N. ....I..K .
	CP	2	.....T...Q.T. PVV..... N..... .R.D.N. ....I..K .
D	BS	c	.....N...Q.T. .V..... N...S... .A...T.....Y .....I.....
	RP	c	.....N...Q.T. .V..... N...S... .T...T..... .....R...I...D N
	RT	c	.....N...Q.T. .V..... N...S... .T...T..... .....R...I...D N
	LT	c	.....N...Q.T. .V..... N...S... .T...T..... .....R...I...D N
	RF	16	V.N...Q.T. .V..... N...S... .T...T..... .I...T...Q...I.....
	LF	12	.....N...Q.T. .V..... N...S... .T...T..... .....R...I...D N
	CP	3	.....N...Q... .V...I... N...A...T..... .....I...VK.....I.....
	RO	c	.....N...Q... .V...I... N...A...T..... .....I...K.....I.....
	LO	c	.....N...Q... .V...I... N...A...T..... .....I...K.....I.....

## C) V4

392

445

A	LN	2	STQLFNSTWD NGEVL--SNT EGNGTTLPC RIKQIINMWQ EVGKAMYAPP IR
B	LN	3	.....P.GT--NI. ....DNI.....L.....
	CP	3	.....RP.GT--NI. ....DNI.....
C	LN	4	.....N GT.GTT-N.. GE--NI.....V.....
	CP	7	.....G.N GT.GTT-N.. GE--NI.....
	BS	8	.....N GT.GTT-N.. GE--NI.....
	LF	19	.....N GT.GTT-N.. GE--NI.....
	RP	10	.....N GT.GTT-N.. GE--NI.....
	LT	10	.....N DT.GTT-N.. GE--NI.....
	RT	10	.....N GT.GTT-N.. GE--NI.....
D	LF	9	.....N DTDIKLN.. ....NI.....
	RF	10	.....N DTDIKLN.. ....NI.....
E	LO	6	.....N GT...--N.. .R.DNI.....
	RO	9	.....N GT...--N.. .R.DNI.....
F	BS	2	.....N DTDERTNI. ....NI.....

## D) V5

446

478

A	LN	4	GQISCSNIT	GIMLTRDGGN	TT-NETTEIF	RPGGG
	CP	4	...R.....	...A...V.....		
B	LN	5	...R.....	.MI.....I	N.RDNK.....	
	CP	6	...R.....	.MI.....I	N.RDNK.....	
	BS	10	...R.....	.MI.....I	N.RDNK.....	
	LF	12	...R.....	.MI.....I	N.RDNK.....	
	RP	10	...R.....	.MI.....I	N.RDNK.....	
	LT	10	...R.....	.MI.....I	N.RDNK.....	
	RT	10	...R.....	.MI.....I	N.RDNK.....	
	C	LF	16	...R.....	.MI.....I	SKS.NDS.V.....
RF		11	...R.....	.MI.....I	SKS.NDS.V.....	
D	LO	4	...R.....	.MI.....I	N.GDNE..V..V..	
	RO	5	...R.....	.MI.....I	N.GDNE..V..V..	

## E) gp41/nef

840

870 1

51

A	LN2	c	RAFRVHIHIP	RRIRQGFERA	LL*LC*DG*	MGNKWSKSSR	SGWDTVRERM	RRAEPAAEPA	AAGVGAASRD	LERHGAISSN
	LN1	c	.....	.....	.....*	.....	.....	.....	.....D	.....
	CP	3	.....	.....	.....*	.....E.....	.....	.....	.....	.....
B	RP	c	.V.....	.....	.....*	RR.PA.....	.....	.....	.....D	.....
	RF	c	.V.....	.....	.....*	N.PA.....	.....	.....	.....D	.....
	LF	c	.V.....	.....	.....*	T.PA.....	.....	.....	.....D	.....X
	LT	c	.V.....	.....	.....*	T.PA.....	.....	.....	.....D	.....
	CP	3	.G.....	.....	.....*	T.PA.....	.....	.....	.....D	.....M
C	RO	c	.....L.....	.....	.....--E..W*	.....G.....	M.PA.....	.....	.....D	.....
	LO	c	.....L.....	.....	.....--E..W*	.....G.....	M.PA.....	.....	.....D	.....
	RT	c	.....L.....	.....	.....--E..W*	.....G.....	M.PA.....	.....	.....D	.....
	BS	c	.....L.....	.....	.....--E..W*	.....G.....	M.PA.....	.....	.....D	.....

FIG. 4. Comparison of inferred amino acid sequences of variants from different regions of the brain of NA234 in V1/V2 (A), V3 (B), V4 (C), V5 (D), and gp41 (E) /nef, using LN variants as reference sequences. Horizontal lines divide primary bootstrap-supported ( $\geq 75\%$  of data sets) phylogenetic groupings of nucleotide sequences. Each number in the third column indicates the number of clones used to create the indicated consensus sequence. c, consensus sequence obtained by direct sequencing of PCR product. Symbols: -, sequence identity with LN sequence; -, gap introduced to presence sequence alignment; \*, termination codon. Sequences are numbered by their positions in the HIV<sub>LAI</sub> gp120 (A to D) or nef (E) sequence. BS, brain stem.



TABLE 2. p17<sup>gag</sup> and V3 sequence divergence

Comparison	Study subject	Distance <sup>a</sup>					
		p17 <sup>gag</sup>			V3		
		Mean	Error	Rank <sup>b</sup>	Mean	Error	Rank
Within brain	NA021	0.034	0.009	3	0.015	0.008	4
	NA173	0.035	0.010	2	0.017	0.008	3
	NA128	0.012	0.006	4	0.035	0.011	1
	NA234	0.042	0.011	1	0.028	0.010	2
Brain-lymphoid tissue	NA021	0.040	0.009	3	0.043	0.014	3
	NA173	0.040	0.011	2	0.031	0.011	4
	NA128	0.023	0.008	4	0.080	0.018	1
	NA234	0.057	0.013	1	0.072	0.018	2

<sup>a</sup> Mean J-C distance between variants within brain or between brain and LN or spleen (excluding CP).<sup>b</sup> Rank of variability (1 = most variable).

genomic regions. To investigate further the differentiation between lymphoid and brain-derived variants, sequences from the p17<sup>gag</sup> and V3 regions were obtained from 8 to 12 brain regions from three other individuals (NA021, NA173, and NA128) and compared with those amplified from the corresponding LN or spleen samples (Table 2; Fig. 5 and 6). Se-

quences from each individual were monophyletic in both p17<sup>gag</sup> and V3 upon comparison with each other and with epidemiologically unlinked subtype B sequences (listed in Materials and Methods).

In each of the three study subjects, p17<sup>gag</sup> sequences of variants infecting different regions of the brain and lymphoid

## A) NA173

		276							357	
A	LN2	VIRSANFTNN	AKVIVQLNK	SVEINCTRPN	NNTRKGIHIG	PGRAFYRTGE	IIGNIRQAHC	DLSKAAWNDT	LKQIVIKLRE	Q
	LN3	.....	.....	I.....	.....	.....	.....	N..GT.....	.....	.....
	BG	.....	.....	.....	.....	A.....	.....	N..R.....	.....	.....
	BS	.....	.....	.....	.....	A.....	.....	N..R.....	.....	.....
	LF	.....	.....	.....	S.....	T.....	.....	N..RT.....	.....	.....
	RF	.....	.....	.....	S.....	.....	.....	N.....	.....	.....
	LT	.....	.....	I.....	.....	A.....	.....	N.....	.....	.....
	RT	.....	.....	.....	.....	A.....	.....	N.....	.....	.....
	RO	.....	.....	E.....	.....	A.....	.....	N..G.....	.....	.....
	RP	.....	.....	.....	.....	A.....	.....	N.....	.....	.....

## B) NA021

276

357

A	LN1 LN2	VIRSENFTNN AKVIVQLNE TVEINCTRPN NNTRKSITMG PGKAFYATGD IIGDIRQAHC NISRTRWNNT LKQIVIKLRE Q																																																																																																																																		
B	LC RC BS LF RF RP BG CP RO RT LP LT LO	<table><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>.....</td><td>.....</td><td>.....</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>N</td><td>.....</td><td>K</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>.....</td><td>.....</td><td>K</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>.....</td><td>.....</td><td>K</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>.....</td><td>.....</td><td>K</td><td>.....</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>.....</td><td>.....</td><td>K</td><td>K</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>.....</td><td>.....</td><td>K</td><td>K</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>.....</td><td>.....</td><td>K</td><td>K</td></tr><tr><td>.....</td><td>N</td><td>.....</td><td>.....</td><td>P</td><td>.....</td><td>.....</td><td>.....</td><td>K</td><td>K</td></tr></table>	.....	N	.....	.....	P	.....	N	.....	.....	.....	.....	N	.....	.....	P	.....	N	.....	.....	.....	.....	N	.....	.....	P	.....	N	.....	.....	.....	.....	N	.....	.....	P	.....	N	.....	.....	.....	.....	N	.....	.....	P	.....	.....	.....	.....	.....	.....	N	.....	.....	P	.....	N	.....	K	.....	.....	N	.....	.....	P	.....	.....	.....	K	.....	.....	N	.....	.....	P	.....	.....	.....	K	.....	.....	N	.....	.....	P	.....	.....	.....	K	.....	.....	N	.....	.....	P	.....	.....	.....	K	K	.....	N	.....	.....	P	.....	.....	.....	K	K	.....	N	.....	.....	P	.....	.....	.....	K	K	.....	N	.....	.....	P	.....	.....	.....	K	K
.....	N	.....	.....	P	.....	N	.....	.....	.....																																																																																																																											
.....	N	.....	.....	P	.....	N	.....	.....	.....																																																																																																																											
.....	N	.....	.....	P	.....	N	.....	.....	.....																																																																																																																											
.....	N	.....	.....	P	.....	N	.....	.....	.....																																																																																																																											
.....	N	.....	.....	P	.....	.....	.....	.....	.....																																																																																																																											
.....	N	.....	.....	P	.....	N	.....	K	.....																																																																																																																											
.....	N	.....	.....	P	.....	.....	.....	K	.....																																																																																																																											
.....	N	.....	.....	P	.....	.....	.....	K	.....																																																																																																																											
.....	N	.....	.....	P	.....	.....	.....	K	.....																																																																																																																											
.....	N	.....	.....	P	.....	.....	.....	K	K																																																																																																																											
.....	N	.....	.....	P	.....	.....	.....	K	K																																																																																																																											
.....	N	.....	.....	P	.....	.....	.....	K	K																																																																																																																											
.....	N	.....	.....	P	.....	.....	.....	K	K																																																																																																																											

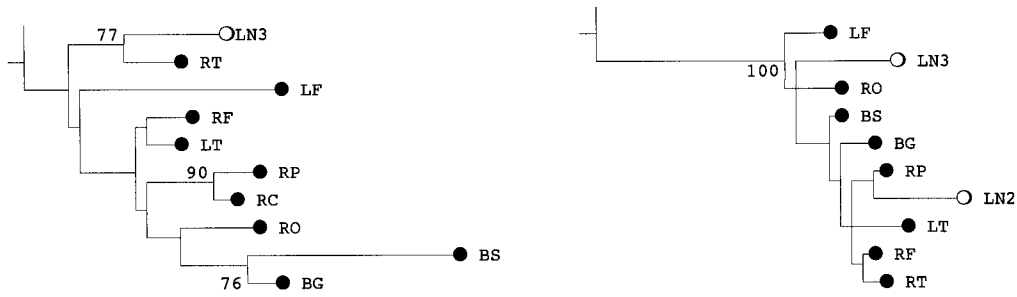
## C) NA128

		276											357
A	SPL RF LF	VIRSENFSDN ..... ..... T.....	AKTIIVHLNE ..... ..... .....	SVEINCTRPN ..... ..... .....	NNTRKRITMG ..... S.P..... S.P.....	PGRVYYTTGQ ..... AL..... AF.....	IIGNIRQAHC ..... ..... .....	NISRERWNKT ..... ..... .....	LEQIVIKLGE ..... ..... .....	K..... ..... .....	K..... ..... .....		
B	RT	.....	N.....	.....	K.....	.....	S.PI.....	AL.....	.....	KG.....	.....		
	LT	.....	N.....	.....	.....	.....	S.PI.....	AI.....	.....	AK.....	R.....		
	LO	.....	N.....	.....	I.....	.....	S.PI.....	AI.....	D.....	AK.....	A.....		
	RP	.....	N.....	V.....	.....	.....	S.PI.....	AI.....	D.....	AK.....	A.....		
	LP	.....	N.....	V.....	.....	.....	S.PI.....	AI.....	D.....	AK.....	A.....		
	BG	.....	N.....	V.....	.....	.....	S.PI.....	AI.....	D.....	AK.....	A.....		
	BS	.....	N.....	.....	.....	.....	S.PI.....	AI.....	D.....	AK.....	A.....		

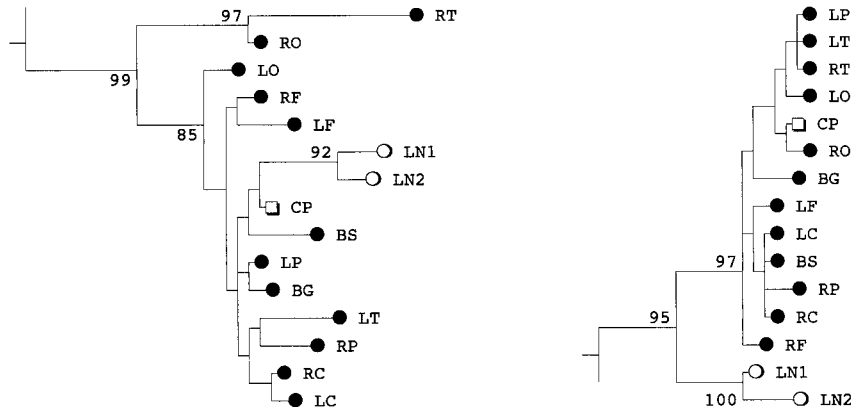
FIG. 5. Comparison of inferred amino acid sequences of variants from the V3 region of NA173 (A), NA021 (B), and NA128 (C), using the LN or spleen (SPL) sequences as a reference. Horizontal lines divide bootstrap-supported ( $\geq 75\%$  of data sets) phylogenetic groupings of nucleotide sequences (symbols and sequence numbering correspond to those in Fig. 4). BS, brain stem.



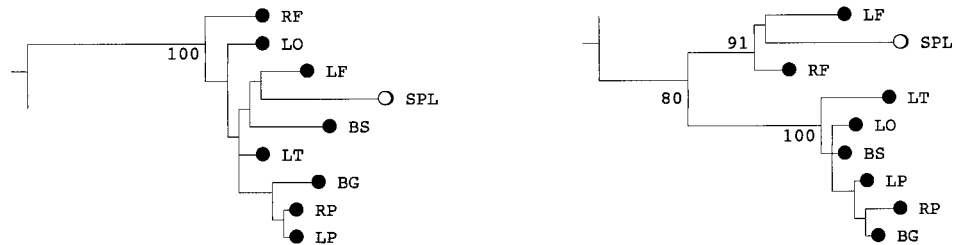
A) NA173



B) NA021



C) NA128



D) NA234

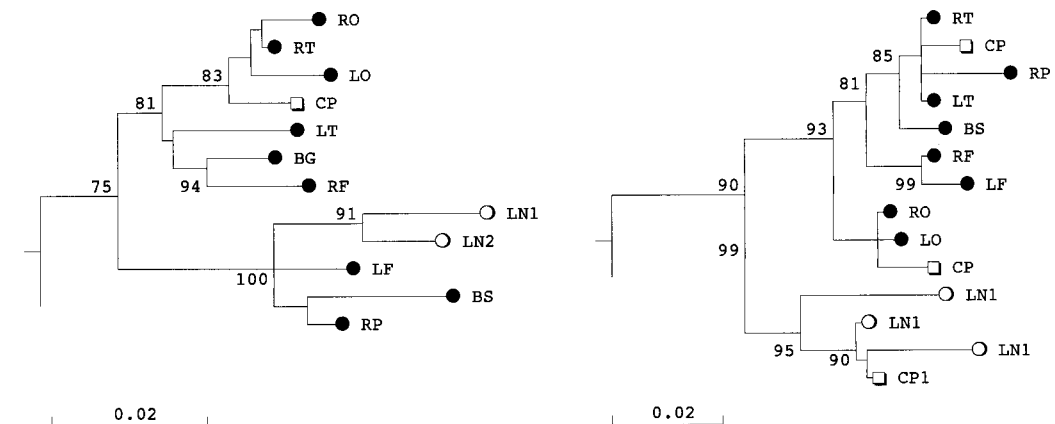


FIG. 6. Phylogenetic analysis of p17<sup>tag</sup> (column 1) and V3 sequences (column 2) of the four study subjects (A through D). Sequences from each of the four subjects were monophyletic in both genomic regions upon comparison with each other and the sequences available from GenBank listed in Materials and Methods and were used to root each clade. p17<sup>tag</sup> and V3 trees from different study subjects were plotted by using the indicated scale of J-C distances. Sequences were derived from brain (●), from lymphoid tissue (LN or spleen [SPL]) (○), or from CP (□). All bootstrap values of  $\geq 75\%$  are indicated on branches. Sequences from NA234 were obtained from individual clones of amplified DNA instead of by direct sequencing; single representative clones from each sample or multiple clones for those containing sequences in more than one lineage (i.e., LN and CP) have therefore been included. BS, brain stem; BG, basal ganglia.

tissue showed interrelationships different from those observed in V3. First, there was no correlation between sequence diversity between the two genomic regions (Table 2). For example, p17<sup>gag</sup> sequences from NA173 showed two- to threefold-greater variability between brain regions (mean J-C distance of  $0.035 \pm 0.010$ ) and between brain and lymphoid tissue (mean J-C distance of  $0.04 \pm 0.011$ ) than the corresponding sequences from NA128 (distances of  $0.012 \pm 0.006$  and  $0.023 \pm 0.008$ ). In contrast, sequences from NA128 in the V3 region were more variable than those from NA173 ( $0.035 \pm 0.011$  and  $0.080 \pm 0.018$ , compared with mean distances of  $0.017 \pm 0.008$  and  $0.031 \pm 0.011$  for NA173).

Second, phylogenetic analysis of the p17<sup>gag</sup> and V3 sequences produced different groupings of variants (Fig. 6). For example, LN sequences from NA021 clustered separately from brain-derived variants in the V3 region but grouped with brain variants in p17<sup>gag</sup>. Conversely, LN sequences from NA173 were distinct from all but one of the variants in p17<sup>gag</sup> but were undifferentiated from brain variants in V3. In each of the four study subjects, tissue specificity depended on the region being compared. A specific instance of incompatible phylogeny comparable to that observed for NA234 (see above) was observed in NA021, where LN sequences formed a bootstrap supported lineage in V3 distinct from the brain variants, while the primary branching order in p17<sup>gag</sup> divided sequences from the RT and RO brain regions from the rest of the sequences.

**Sequence divergence and pathology appearance.** The severity of HIV-1 varied between the study subjects (Table 1), ranging from infrequent giant cells confined to perivascular regions (NA173) to widespread pathology affecting both white and grey matter (NA128; Fig. 7). The extent and type of HIV-induced pathology correlated with the degree of V3 but not p17<sup>gag</sup> sequence diversity between different brain regions and in the extent to which V3 sequences from the brain differed from those in lymphoid tissue. V3 sequences from NA173 (who showed minimal HIV-related pathology) were largely undifferentiated from those detected in LN, with only an alanine-arginine change segregating by tissue (mean J-C distances listed in Table 2). At the other extreme, the spleen-derived sequence from NA128 had a predicted syncytium-inducing phenotype and differed at multiple sites from the nonsyncytium-inducing variants in the brain. In the p17<sup>gag</sup> region, however, sequences from NA173 were the most variable, and those from NA128 were the least variable.

## DISCUSSION

This study documents the extraordinary complexity of HIV populations *in vivo*, findings which have several implications in understanding mechanisms of tissue adaptation and the timing of entry of HIV into the central nervous system (CNS). The marked sequence diversity of the p17<sup>gag</sup> region and the readily identifiable clusters of sequences in the hypervariable regions of *env* (V1/V2, V3, V4, and V5) of NA234 and other study subjects provided evidence for recombination between different regions of the genome. For example, variants infecting separate regions of the brain of NA234 were assembled from two different p17<sup>gag</sup> lineages and a limited number of distinct hypervariable region lineages, often with different combinations within same autopsy sample (e.g., the V4 and V5 sequences in the LN, CP, and LF sequences [Fig. 3 to D]). In the other three study subjects, the different degrees of variability and the discordant phylogenetic groupings between p17<sup>gag</sup> and V3 regions indicate a lack of genetic linkage between these two subgenomic regions and further support the hypothesis of frequent recombination *in vivo*.

To date, recombination has been most easily identified between different subtypes of HIV-1; for example, variants of HIV-1 from Thailand contain *gag* sequences resembling those of subtype A but distinct from subtype A in the *env* gene (7), while other viruses appear to have been generated by multiple recombination events (e.g., HIV-1<sub>MAL</sub> [10, 37] and subtype I [21]). Recombination has also been observed upon infection with different strains of HIV-1, either experimentally in a chimpanzee exposed to the laboratory isolates HIV<sub>LAI</sub> and HIV-1<sub>SF2</sub> (47) or possibly through multiple exposure to two or more sources of HIV infection in a blood recipient and an injecting drug user (12, 49). We have now demonstrated that recombination also occurs within an infected individual between variants descended in each case from the original infecting strain. The finding that different parts of the HIV genome can have different evolutionary histories severely limits the concept of tissue specificity of variants of HIV *in vivo*, particularly if these conclusions are based on a single subgenomic region. For example, the brain-specific and frequently monophyletic nature of HIV sequences in the *pol* region (virodemes) of variants infecting antiviral agent-treated individuals (48) may not be reflected elsewhere in the genome. Indeed, sequence relationships in the *env* gene may differ substantially from *pol*, as variation in the former region is more likely to confer phenotypic differences in cellular tropism. The existence of recombination provides an explanation for discordant phylogenies between p17<sup>gag</sup>, V1/V2, and V3 that we observed between brain (in the LF region) and LN sequences in three previous subjects (14, 23, 24). While sequences in V3 were tissue specific, sequences in the p17<sup>gag</sup> and V1/V2 regions were diverse, and some evolutionary lineages were common to variants recovered from brain, lung, and LN. Our observations support the hypothesis that recombination may accompany the acquisition of antiviral resistance, as exemplified by the appearance of zidovudine-resistant mutants in the peripheral circulation which occurred without evidence for a comparable bottlenecking in *env* (6); V3 sequences showed no reduction in diversity during the process of population replacement in *pol* region.

The diversity of V3 sequences in different brain regions of the four study subjects was similar to a previous comparison of variants infecting different brain regions (mean pairwise distance between brain regions, 0.021 [8], excluding sequences from the LF region that were highly divergent in sequence and failed to group phylogenetically with sequences derived from other regions of the brain). The LF sequences may have originated from exogenous contamination of the PCR, or corresponded to an epidemiologically unlinked isolate in a case of mixed infection. In either case, the observed degree of sequence divergence was unlikely to have originated from sequence change over the course of infection within the study subject.

Greater degrees of sequence complexity may also originate from the presence of different infected cell types in a tissue sample. HIV sequences amplified from the choroid plexus of NA234 showed the greatest diversity in the *env* region, containing variants corresponding to those from lymphoid tissue and brain, consistent with the presence of virus from blood-derived cells and brain parenchyma. The proximity of these different cell types in the CP may provide an opportunity for recombination of HIV to occur, as well as a site of entry of HIV into the CNS. Without biological characterization of the variants found in the CP or elsewhere in the brain, it remains unclear whether recombinant genomes have been selected or represent random samplings of phenotypically identical viruses. However, the multiple recombination events observed in

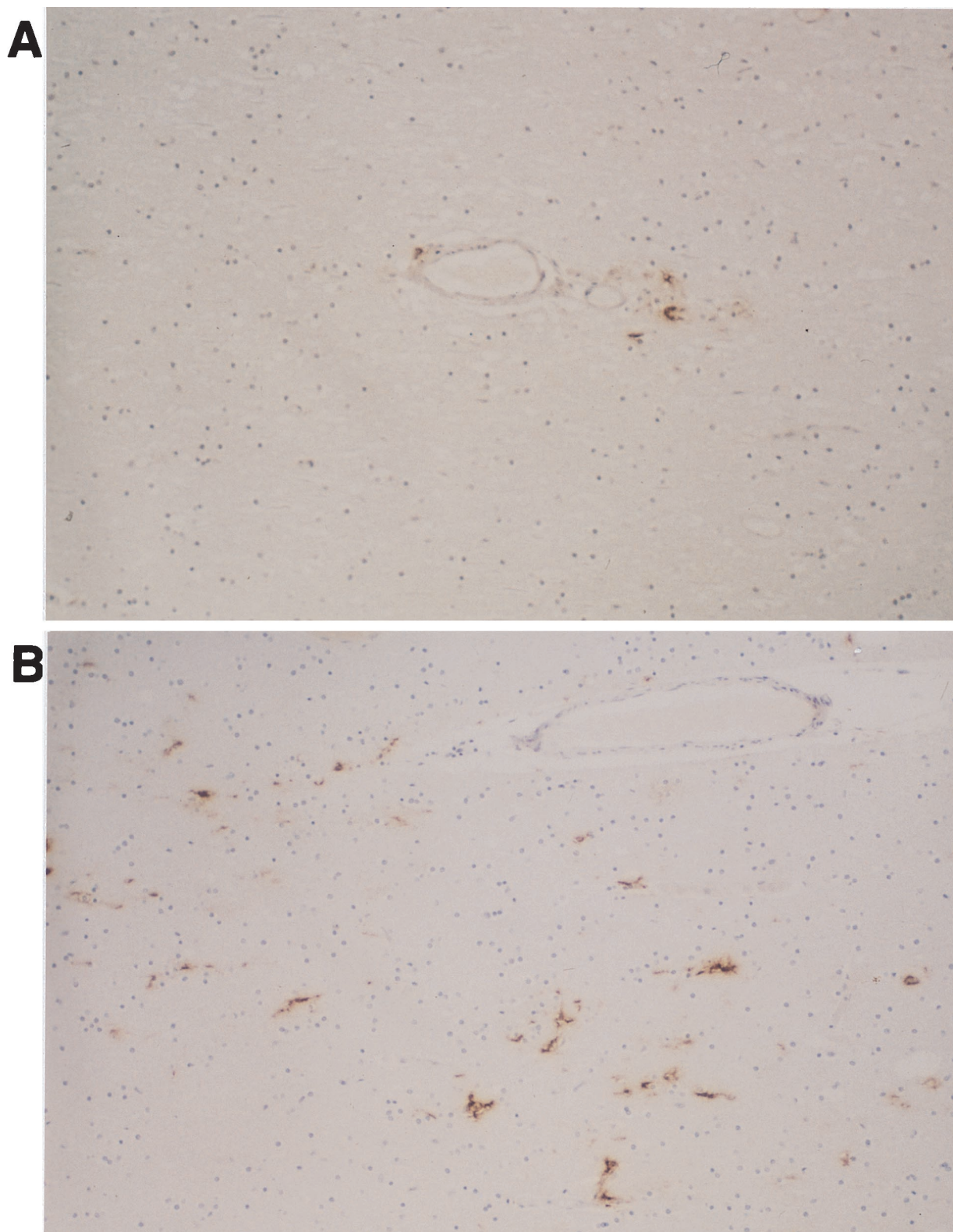


FIG. 7. Immunocytochemical detection of p24 antigen in representative sections of cerebral white matter of two study subjects with HIV. (A) NA173. Immunopositive mononuclear and giant cells (stained brown with diaminobenzidine) are confined to the perivascular region. (B) NA128. Widely dispersed HIV-infected microglia in white matter.



this study would provide a powerful mechanism for adaptation, providing, for example, an effective method for the spread of antiviral agent resistance or cytotoxic T-cell escape mutants into the CNS. Recombination between these latter determinants and *env* could produce new virus populations that retain their neuroadapted phenotype.

The differing sequence relationships between brain-derived and lymphoid variants from the four study subjects suggests that entry of HIV-1 into the CNS can occur at different times. The current consensus view that entry occurs early during HIV infection is supported by the observation that HIV RNA sequences can be detected in cerebrospinal fluid throughout the course of infection and by the detection of low levels of HIV proviral sequences in brains of asymptomatic individuals (4, 15, 41). Early entry is also supported by the extensive sequence diversity in the p17<sup>gag</sup> region of variants recovered from the brain, such as between lineages A and B observed in NA234 in this study, which implies several years of divergent evolution (24). However, the relevance of early entry into the CNS in the development of late stage HIV remains unclear, since active virus replication has not been demonstrated immunocytochemically during early infection (4), and brains show little evidence of pathology apart from the presence of infiltrating CD8 lymphocytes in perivascular areas.

Evidence for a contribution of late-entering variants to HIV is provided by NA173, who showed a distinctive pathology appearance of HIV-expressing infiltrating macrophages confined to the perivascular regions (Table 1; Fig. 7A). The hypothesis of recent entry of HIV-infected cells into the brain parenchyma was supported by the observation of close sequence similarity in the V3 region of brain-derived variants with those obtained from lymphoid cells. This late-entry picture contrasted strongly with the distribution of HIV infection in NA128, in which HIV was widely dispersed in white (Fig. 7B) and grey matter, while V3 sequences were distinct between spleen and brain and heterogeneous within brain (Table 2; Fig. 5C). This correlation was, however, not supported by sequence comparisons in the p17<sup>gag</sup> region, where sequence diversity was greatest in NA173 and least in NA128.

Indeed, to understand the adaptive significance of the sequence differences in different parts of the genome, it will in the future be necessary to analyze functionally the contribution of each genomic region to the phenotype of the virus. In particular, it will be important to determine the phenotypic significance of recombination between the p17<sup>gag</sup> and *env* regions, particularly as variants with different combinations of lineages in the two regions were associated with distinct pathology appearances. Understanding what contributes to neurotropism will illuminate the selective pressures (if any) that produce the recombinant viruses observed in this study. The lack of genetic linkage in the HIV genome resulting from recombination greatly enhances its ability to adapt to several simultaneously acting selection pressures, as indicated by the rapid emergence in vitro of dual antiviral agent-resistant mutants (32).

#### ACKNOWLEDGMENTS

We are grateful to Francis Brannan for preparation of the sections for immunocytochemistry and for the provision of frozen samples from the MRC Edinburgh Brain Bank. We also greatly appreciate the critical review and discussion contributed by Donald Smith.

This study was supported by Medical Research Council Strategic Project grant G9632414.

#### REFERENCES

- Alkhatib, G., C. Combadiere, C. C. Broder, Y. Feng, P. E. Kennedy, P. M. Murphy, and E. A. Berger. 1996. CC CKR5: a RANTES, MIP-1 alpha, MIP-1 beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**:1955-1958.
- Asjo, B., J. Albert, A. Karlsson, L. Morfeldt Manson, G. Biberfeld, K. Lidman, and E. M. Fenyo. 1986. Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* **ii**:660-662.
- Ball, J. K., E. C. Holmes, H. Whitwell, and U. Desselberger. 1994. Genomic variation of human immunodeficiency virus type 1 (HIV-1)—molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy. *J. Gen. Virol.* **75**:867-879.
- Bell, J. E., A. Busuttil, J. W. Ironside, S. Rebus, Y. K. Donaldson, P. Simmonds, and J. F. Peutherer. 1993. Human immunodeficiency virus and the brain—investigation of virus load and neuropathologic changes in pre-AIDS subjects. *J. Infect. Dis.* **168**:818-824.
- Berson, J. F., D. Long, B. J. Doranz, J. Rucker, F. R. Jirik, and R. W. Doms. 1996. A seven-transmembrane domain receptor involved in fusion and entry of T-cell-tropic human immunodeficiency virus type 1 strains. *J. Virol.* **70**:6288-6295.
- Brown, A. J. L., and A. Cleland. 1996. Independent evolution of the *env* and *pol* genes of HIV-1 during zidovudine therapy. *AIDS* **10**:1067-1073.
- Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Arstenstein, P. A. Hegerich, D. St. Louis, D. S. Burke, and F. E. McCutchan. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* **70**:5935-5943.
- Chang, J., R. Jozwiak, B. Wang, T. Ng, Y. C. Ge, W. Bolton, D. E. Dwyer, C. Randle, R. Osborn, A. L. Cunningham, and N. K. Saxena. 1998. Unique HIV type 1 V3 region sequences derived from six different regions of brain: region-specific evolution within host-determined quasispecies—short communication. *AIDS Res. Hum. Retroviruses* **14**:25-30.
- Cheng-Mayer, C., D. Seto, M. Tateno, and J. A. Levy. 1988. Biological features of HIV-1 that correlate with virulence in the host. *Science* **240**:80-82.
- Cornelissen, M., G. Kampinga, F. Zorgdrager, and J. Goudsmit. 1996. Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes. *J. Virol.* **70**:8209-8212.
- Deng, H. K., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Dimarzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman, and N. R. Landau. 1996. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**:661-666.
- Diaz, R. S., E. C. Sabino, A. Mayer, J. W. Mosley, and M. P. Busch. 1995. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. *J. Virol.* **69**:3273-3281.
- Di Stefano, M., S. Wilt, F. Gray, M. Dubois-Dalcq, and F. Chiodi. 1996. HIV type 1 V3 sequences and the development of dementia during AIDS. *AIDS Res. Hum. Retroviruses* **12**:471-476.
- Donaldson, Y. K., J. E. Bell, E. C. Holmes, E. S. Hughes, H. K. Brown, and P. Simmonds. 1994. In vivo distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. *J. Virol.* **68**:5991-6005.
- Donaldson, Y. K., J. E. Bell, J. W. Ironside, R. P. Brettell, J. R. Robertson, A. Busuttil, and P. Simmonds. 1994. Redistribution of HIV outside the lymphoid system with onset of AIDS. *Lancet* **343**:382-385.
- Dragic, T., V. Litwin, G. P. Allaway, S. R. Martin, Y. X. Huang, K. A. Nagashima, C. Cayanan, P. J. Maddon, R. A. Koup, J. P. Moore, and W. A. Paxton. 1996. HIV-1 entry into CD4(+) cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**:667-673.
- Epstein, L. G., C. Kuiken, B. M. Blumberg, S. Hartman, L. R. Sharer, M. Clement, and J. Goudsmit. 1991. HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. *Virology* **180**:583-590.
- Felsenstein, J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5**:164-166.
- Feng, Y., C. C. Broder, P. E. Kennedy, and E. A. Berger. 1996. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**:872-877.
- Fenyo, E. M., L. Morfeldt Manson, F. Chiodi, B. Lind, A. von Gegerfelt, J. Albert, E. Olsson, and B. Asjo. 1988. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J. Virol.* **62**:4414-4419.
- Gao, F., D. L. Robertson, C. D. Carruthers, Y. Y. Li, E. Bailes, L. G. Kostrikis, M. O. Salminen, F. Bibollettruche, M. Peeters, D. D. Ho, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1998. An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol.* **72**:10234-10241.
- Haggerty, S., and M. Stevenson. 1991. Predominance of distinct viral genotypes in brain and lymph node compartments of HIV-infected individuals. *Viral Immunol.* **4**:123-131.
- Hughes, E. S., J. E. Bell, and P. Simmonds. 1997. Investigation of population diversity of human immunodeficiency virus type 1 in vivo by nucleotide sequencing and length polymorphism analysis of the V1/V2 hypervariable region of *env*. *J. Gen. Virol.* **78**:2871-2882.

24. Hughes, E. S., J. E. Bell, and P. Simmonds. 1997. Investigation of the dynamics of the spread of human immunodeficiency virus to brain and other tissues by evolutionary analysis of sequences from the p17<sup>gag</sup> and *env* genes. *J. Virol.* **71**:1272–1280.
25. Kasper, P., P. Simmonds, K. E. Schneweis, R. Kaiser, B. Matz, J. Oldenburg, H. H. Brackmann, and E. C. Holmes. 1995. The genetic diversification of the HIV type 1 gag p17 gene in patients infected from a common source. *AIDS Res. Hum. Retroviruses* **11**:1197–1201.
26. Keys, B., J. Karis, B. Fadeel, A. Valentin, G. Norkrans, L. Hagberg, and F. Chiodi. 1993. V3 sequences of paired HIV-1 isolates from blood and cerebrospinal fluid cluster according to host and show variation related to the clinical stage of disease. *Virology* **196**:475–483.
27. Korber, B. T. M., K. J. Kunstman, B. K. Patterson, M. Furtado, M. M. McEvilly, R. Levy, and S. M. Wolinsky. 1994. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *J. Virol.* **68**:7467–7481.
28. Kozak, S. L., E. J. Platt, N. Madani, F. E. Ferro, K. Peden, and D. Kabat. 1997. CD4, CXCR-4, and CCR-5 dependencies for infections by primary and laboratory-adapted isolates of human immunodeficiency virus type 1. *J. Virol.* **71**:873–882.
29. Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, Pa.
30. Li, Y. X., J. C. Kappes, J. A. Conway, R. W. Price, G. M. Shaw, and B. H. Hahn. 1991. Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes. *J. Virol.* **65**:3973–3985.
31. Liu, Z. Q., C. Wood, J. A. Levy, and C. Cheng Mayer. 1990. The viral envelope gene is involved in macrophage tropism of a human immunodeficiency virus type 1 strain isolated from brain tissue. *J. Virol.* **64**:6148–6153.
32. Moutouh, L., J. Corbeil, and D. D. Richman. 1996. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc. Natl. Acad. Sci. USA* **93**:6106–6111.
33. Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
34. Pang, S., H. V. Vinters, T. Akashi, W. A. O'Brien, and I. S. Chen. 1991. HIV-1 *env* sequence variation in brain tissue of patients with AIDS-related neurologic disease. *J. Acquir. Immune. Defic. Syndr.* **4**:1082–1092.
35. Power, C., J. C. McArthur, R. T. Johnson, D. E. Griffin, J. D. Glass, S. Perryman, and B. Chesebro. 1994. Demented and nondemented patients with AIDS differ in brain-derived human immunodeficiency virus type 1 envelope sequences. *J. Virol.* **68**:4643–4649.
36. Reddy, R. T., C. L. Achim, D. A. Sirko, S. Tehranchi, F. G. Kraus, F. Wong-Staal, C. A. Wiley, I. Grant, J. H. Atkinson, M. Kelly, J. L. Chandler, M. R. Wallace, J. A. McCutchan, S. A. Spector, L. Thal, R. K. Heaton, J. Hesselink, T. Jernigan, E. Masliah, I. Abramson, N. Butters, T. Patterson, S. Zisook, and D. Jeste. 1996. Sequence analysis of the V3 loop in brain and spleen of patients with HIV encephalitis. *AIDS Res. Hum. Retroviruses* **12**:477–482.
37. Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn. 1995. Recombination of HIV-1. *Nature* **374**:124–126.
38. Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown. 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **64**:5840–5850.
39. Simmonds, P., P. Balfe, J. F. Peutherer, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown. 1990. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J. Virol.* **64**:864–872.
40. Simmonds, G., D. Wilkinson, J. D. Reeves, M. Dittmar, S. Beddows, J. Weber, G. Carnegie, U. Desselberger, P. W. Gray, R. A. Weiss, and P. R. Clapham. 1996. Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J. Virol.* **70**:8355–8360.
41. Sinclair, E., F. Gray, and F. Scaravilli. 1992. PCR detection of HIV proviral DNA in the brain of an asymptomatic HIV-positive patient. *J. Neurol.* **239**:469–471.
42. Steuler, H., B. Storch Hagenlocher, and B. Wildemann. 1992. Distinct populations of human immunodeficiency virus type 1 in blood and cerebrospinal fluid. *AIDS Res. Hum. Retroviruses* **8**:53–59.
- 42a. Strappe, P. Personal communication.
43. Stuyver, L., A. Wyseur, A. Rombout, T. Scarcez, C. Verhofstede, D. Rimland, R. F. Schinazi, and R. Rossau. 1997. Line probe assay for rapid detection of drug-selected mutations in the human immunodeficiency virus type 1 reverse transcriptase gene. *Antimicrob. Agents Chemother.* **41**:284–291.
44. Tersmette, M., J. M. Lange, R. E. de Goede, F. de Wolf, J. K. Eeftink Schattenkerk, P. T. Schellekens, R. A. Coutinho, J. G. Huisman, J. Goudsmit, and F. Miedema. 1989. Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality. *Lancet* **i**:983–985.
45. Vanderhoek, L., C. J. A. Sol, J. Maas, V. V. Lukashov, C. L. Kuiken, and J. Goudsmit. 1998. Genetic differences between human immunodeficiency virus type 1 subpopulations in faeces and serum. *J. Gen. Virol.* **79**:259–267.
46. Vanderhoek, L., C. J. A. Sol, F. Snijders, J. F. W. Bartelsman, R. Boom, and J. Goudsmit. 1996. Human immunodeficiency virus type 1 RNA populations in faeces with higher homology to intestinal populations than to blood populations. *J. Gen. Virol.* **77**:2415–2425.
47. Wei, Q., and P. N. Fultz. 1998. Extensive diversification of human immunodeficiency virus type 1 subtype B strains during dual infection of a chimpanzee that progressed to AIDS. *J. Virol.* **72**:3005–3017.
48. Wong, J. K., C. C. Ignacio, F. Torriani, D. Havlir, N. J. S. Fitch, and D. D. Richman. 1997. In vivo compartmentalization of human immunodeficiency virus: evidence from the examination of *pol* sequences from autopsy tissues. *J. Virol.* **71**:2059–2071.
49. Zhu, T. F., N. Wang, A. Carr, S. Wolinsky, and D. D. Ho. 1995. Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype b in an acute seroconverter. *J. Virol.* **69**:1324–1327.